Acoustic characteristics of English diphthongs produced by Japanese learners

Aya Kitagawa

Graduate School of Education, Waseda University

aya@fuji.waseda.jp

Abstract

This study aimed to characterize the English diphthongs produced by the Japanese learners of English, compared to the native speakers of English. The target diphthongs were [eɪ, ɔɪ, aɪ, əʊ/oʊ, aʊ]. The duration, trajectory length and spectral rate of change were measured. It was revealed that [eɪ] was problematic to the Japanese learners.

Keywords

diphthongs, Japanese learners, acoustic analysis

1 Introduction

1.1 The Japanese-learner diphthongs

Japanese language does not have diphthongs because two vowels produced next to one another are recognized to comprise two independent syllables. In spite of the lack of diphthongs in Japanese, there are few studies on how Japanese learners produce English diphthongs. One of the limited numbers of research on this issue was conducted by Nakamura, Suzuki, Minematsu, Hirose and Makino (2010). They mentioned that Japanese learners tend to produce both elements too clearly because each element constitutes a different syllable and is distinctly articulated in Japanese.

1.2 English diphthongs

According to Jones (2003), British English has eight diphthongs [eɪ, aɪ, ɔɪ, əʊ, aʊ, ɪə, eə, ʊə] while American English has three diphthongs [aɪ, ɔɪ, aʊ], called wide diphthongs. There are fewer American English diphthongs because [eɪ, oʊ] ([oʊ] corresponds to British [əʊ]) are classified as tense vowels due to only a little diphthongal movement and the schwa of [ɪə, eə, ʊə] is replaced by [ɹ]. This study dealt with [eɪ, ɔɪ, aɪ, əʊ, aʊ] as English diphthongs.

1.3 Hypotheses

The diphthongal movement is small for [eɪ, əʊ/oʊ] in American English as mentioned above; therefore, the Japanese learners would not produce them as a clear diphthong, either. Furthermore, this hypothesis would be more plausible, considering that diphthongs are often changed to long monophthongs in words loaned to Japanese. The rest of the diphthongs, [ɔɪ, aɪ, aʊ], would be much less difficult for Japanese learners to produce because they consist of the two elements distant from each other, as grouped into wide diphthongs in American English.

2 Methodology

2.1 Participants

Participants were 15 speakers of English: five Japanese learners (JL), five native speakers of American English (AN) and British English (BN) each. The JL participants were high school students in the third grade. The data of the BN and AN participants were provided by the UCL Speaker Database (Markham & Hazan, 2002) and the AUE Audio Archive (Merfert, 1997) respectively.

2.2 Material

A phonetically-balanced passage, “The Story of Arthur the Rat,” was used for the experiment of the JN and BN and “Arthur the Rat” for AN. These two passages slightly differ in some words used.

The target diphthongs included five closing diphthongs, [eɪ, ɔɪ, aɪ, əʊ/oʊ, aʊ] and the tokens containing these target diphthongs were selected from the passages for this study. They were “named, <u>take</u>, make, say, day, came, face, (unsafe)” for [eɪ], “choice, noise, (join, boys)” for [ɔɪ], “mind, <u>sight</u> <u>sighed</u>, night, <u>side</u>, (five, nice)” for [aɪ], “only, know, no, home, stone, don’t, go” for [əʊ/oʊ], and “out, scouts, found, house(s), cow, down, about” [aʊ]. The tokens only in “The Story of Arthur the Rat” are underlined and those only in “Arthur the Rat” are bracketed off.

2.3 Acoustic measurements

The speeches were acoustically analyzed with Praat to obtain the duration and the first formant (F1) and second formant (F2). The F1 and F2 values were especially measured at the 20%, 35%, 50%, 65% and 80% point of each vowel to examine the

dynamic characteristics of diphthongs, normalized using Lobanov's (1971) procedure and converted into mel. Then, the trajectory length (TL) and the spectral rate of change (ROC) were calculated as measures of the diphthongal dynamics. This method was suggested by Fox and Jacewicz (2009).

3 Results and discussion

Figure 1 represents the formant movement of the diphthongs. The F1 and F2 mel values were averaged across the participants for each group.

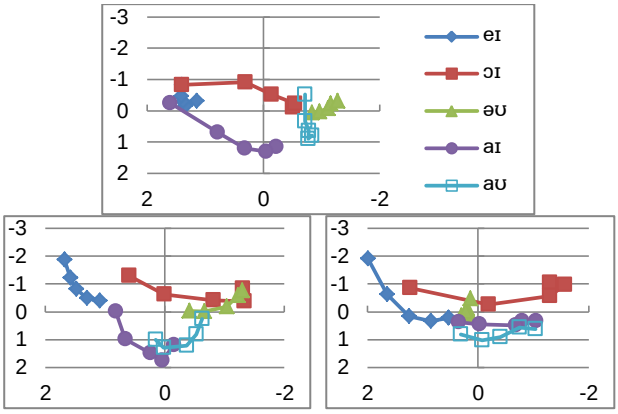

Figure 1: The Formant Movement of JN (top), AN (lower left) and BN (lower right)

A one-way ANOVA was performed for each diphthong and each measure. Due to space limitations, only the diphthongs whose significant difference was yielded between JL and BN/AN for the duration, TL or ROC will be discussed below.

A significant difference was found for [eɪ]'s TL and ROC, $F(2, 12) = 38.8$, $p = .00$. and $F(2, 12) = 14.3$, $p = .001$ respectively, and [ɔɪ]'s TL, $F(2, 12) = 15.6$, $p = .002$. A Turkey's HSD test identified a difference between JL and BN for both TLs and between JL and both native groups for [eɪ]'s ROC. Because JL's TLs of [eɪ, ɔɪ] did not significantly differ from AN at least, the results show that [eɪ]'s ROC was the only problem for JL. Table 1 corroborates JL's fast ROC for [eɪ]. Seemingly, it suggests that JL simply used a very fast manner in articulating this diphthong. However, JL's fast ROC highly likely resulted from the lack of a formant movement as little formant movement is displayed in Figure 1, though the difference between JL and AN did not reach a significant level.

Table 1: The ROC of [eɪ]

	JL		AN		BN	
	M	SD	M	SD	M	SD
eɪ	**287.3**	**103.6**	**521.8**	**68.2**	**745.8**	**169.6**

Thus, the hypotheses were mostly confirmed. First, as hypothesized, the three diphthongs produced by JL, [ɔɪ, aɪ, aʊ], were not significantly different from those by AN and/or BN for all three measures. Second, JL's [əʊ/oʊ] was predicted to be produced in a long-vowel-like manner, but it did not differ significantly from AN's and BN's. AN's [oʊ] showed a small formant movement, too, which possibly led JN to follow their way of articulation. Finally, the only possible problem this study pointed out was JL's production of [eɪ]. JL's [eɪ] had the smallest formant movement, which resulted in the fast ROC. Therefore, the hypothesis on JL's difficulty in producing [eɪ] was proved.

4 Conclusion

It was found out that the Japanese learners produced [ɔɪ, aɪ, əʊ/oʊ, aʊ] successfully while [eɪ] was problematic to them. However, one of the limitations of this study is that it did not compare the location of these diphthongs in a vowel space among the three groups. For instance, there was no significant difference yielded for [aɪ] concerning the three measures applied here, but JL's distribution of [aɪ] appeared very different from AN's and BN's as far as the figures are examined. If the F1 and F2 values are directly compared, the difference between the Japanese learners and the native speakers might be detected.

Acknowledgement

The author expresses her gratitude to the providers of the publicly available speech databases, the ULC Speaker Database and the AUE Audio Archive.

References

Fox, A. L., & Jacewicz, E. (2009) Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *126*(5), 2603-2618.

Jones, D. (2003). *Cambridge English Pronouncing Dictionary*. Cambridge: CUP.

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *Journal of the Acoustical Society of America*, *49*, 606-608.

Markham, D., & Hazan, V. (2002). The UCL speaker database. *Speech, Hearing and Language: UCL Work in Progress*, *14*, 1-17.

Merfert, I. (1997). The alt.usage.english Audio Archive. Retrieved from http://alt-usage-english.org/index.shtml

Nakamura, A., Suzuki, M., Minematsu, N., Hirose, K., & Makino, M. (2010). An experimental study on assessment of diphthongs of learners using pronunciation structure. *Proceedings of the 2010 Spring Meeting of the Acoustical Society of Japan*, *1-P-15*, 439-442.